bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Monod model is insufficient to explain biomass growth in nitrogen-limited yeast fermentation

David Henriques,[a,*] Eva Balsa-Canto[a]

Bioprocess engineering group, IIM-CSIC, Vigo, Spain [a];

Address correspondence to David Henriques, davidh@iim.csic.es

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The yeast *Saccharomyces cerevisiae* is an essential microorganism in food biotechnology; particularly, in wine and beer making. During wine fermentation, yeasts transform sugars present in the grape juice into ethanol and carbon dioxide. The process occurs in batch conditions and is, for the most part, an anaerobic process. Previous studies linked limited-nitrogen conditions with problematic fermentations, with negative consequences for the performance of the process and the quality of the final product. It is, therefore, of the highest interest to anticipate such problems through mathematical models. Here we propose a model to explain fermentations under nitrogen-limited anaerobic conditions. We separated the biomass formation into two phases: growth and carbohydrate accumulation. Growth was modelled using the well-known Monod equation while carbohydrate accumulation was modelled by an empirical function, analogous to a proportional controller activated by the limitation of available nitrogen. We also proposed to formulate the fermentation rate as a function of the total protein content when relevant data are available. The final model was used to successfully explain experiments taken from the literature, performed under normal and nitrogen-limited conditions. Our results revealed that Monod model is insufficient to explain biomass formation kinetics in nitrogen-limited fermentations of *S. cerevisiae*. The goodness-of-fit of the herewith proposed model is superior to that of previously published models, offering the means to predict, and thus control fermentations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Importance: Problematic fermentations still occur in the winemaking industrial practise. Problems include sluggish rates of fermentation, which have been linked to insufficient levels of assimilable nitrogen. Data and relevant models can help anticipate poor fermentation performance. In this work, we proposed a model to predict biomass growth and fermentation rate under nitrogen-limited conditions and tested its performance with previously published experimental data. Our results show that the well-known Monod equation does not suffice to explain biomass formation.

1 Introduction

The yeast species *Saccharomyces cerevisiae* is the best-studied eukaryote. *S. cerevisiae* presents unique characteristics such as its fermentation capacity and its ability to withstand adverse conditions of osmolarity and low pH. These desirable properties have made of *S. cerevisiae* one of the workhorses of bio-industries including food, beverage -especially wine- and biofuel production industries (1).

In winemaking, the fermentation turns grape must into an alcoholic beverage. Wine production occurs in a closed system (i.e. in batch conditions) and is, for the most part, an anaerobic process. During fermentation, yeasts transform sugars present in the must into ethanol and carbon dioxide. A complete alcoholic fermentation is achieved when the residual fermentable sugar is less than 2 g/L. Despite improvements in fermentation control, problematic fermentations such as, stuck - with a higher than desired sugar residual- and sluggish -unusually long- fermentations still occur in real practice.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Various studies have shown that insufficient levels of assimilable nitrogen contribute to stuck or sluggish fermentations (2). A minimum of 120–140 mg/L of assimilable nitrogen is required to achieve a standard fermentation rate, while nitrogen content in grape juice may be as low as 60 mg/L (3). Therefore, it has become common practice to supplement nitrogen-deficient musts with diammonium phosphate (4). Nevertheless, both excessive or insufficient nitrogen can lead to the production of undesired metabolites affecting the organoleptic properties of wine (4, 5).

The need to predict and control wine fermentation has motivated a quest for mechanistic models of alcoholic fermentation (see the reviews by (6, 7)). Previous studies proposed various alternatives that differ on the way they explain biomass formation and the relation of cell mass (or sometimes cell numbers) to fermentation rate. Dynamic kinetic models incorporate the Monod equation to describe biomass growth (8–10). The estimation of the Monod parameters requires nitrogen and sugars uptake data throughout the fermentation. Otherwise, lack of identifiability of the parameters will limit the predictive capability of the model, and thus logistic functions may be more suitable (11, 12). Both formulations, Monod or logistic, achieve a maximum biomass value which can not be exceeded.

However, (13) showed that, for nitrogen-limited anaerobic fermentations of *S. cerevisiae*, there is a substantial increment in biomass content after the depletion of ammonia. Besides, measurements of protein and messenger RNA (mRNA) contents revealed that their concentrations stay stable while the concentration of carbohydrates increases. Similar effects have been detected in baker's yeast exposed to nitrogen starvation but not in the carbon starved cells; see (14) for a fed-batch aerobic example, or (15) for a chemostat anaerobic example. These data would imply that, at least for ammonium deficient scenarios, the use of the Monod model would be insufficient.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Similarly, (16) showed that protein/carbohydrate fractions in the biomass of *S. cerevisiae* vary substantially throughout the fermentation of wine fermentations. To account for this dynamics, (17) and (18) included, in their flux balance analysis models, an empirical function that controls the concentration of carbohydrates in the newly formed biomass as a function of extracellular sugar.

In what regards to the modelling of fermentation rate, previous studies proposed two alternatives. The first implies that the fermentation rate is proportional to biomass or cell numbers (8–10, 12, 19). The second includes the role of hexose transporters as a separate entity that is dependent on the transport of ammonia (11, 20–23).

The advantage of the second over the first is that it enables the simulation of the effect of nitrogen additions. Still, further experimental analysis of the underlying hypothesis is required. Also, a good adjustment to the data comes at the cost of a large number of estimated parameters.

(17) and (18) proposed a third alternative. These authors assumed that cells use different hexose transporters at different fermentation stages based on previously published experimental data (24). However, this approach also requires the adjustment of the maximum transport rate for each hexose carrier. As the maximum transport rate is deemed to be a function of temperature and the initial nitrogen concentration, the modelling also involves the estimation of a substantial number of parameters.

In this work, we propose a dynamic kinetic model capable of successfully explaining nitrogen-limited fermentations. The model describes growth and fermentation rate while distinguishing protein from carbohydrate fractions of biomass. The Monod equation describes the growth, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the carbohydrate accumulation observed after the depletion of ammonia is described by an empirical function, analogous to a proportional controller.

We tested the model properties by fitting the data provided by (13) and (16) who explored fermentations under low and high nitrogen regimes. Our results show why standard Monod models often produce poor results in modelling nitrogen-limited fermentations of *S. cerevisiae*.

Fermentation rate can be a function of the biomass or, as proposed here, a function of the protein fraction since the glycolytic enzymes are an important part of the protein pool. Remarkably, even if protein content is not measured, the model is identifiable provided the uptake of hexoses and nitrogen, as well as the production of ethanol, are measured over time.

We compared the performance of the proposed model with the Monod model (25) and the model proposed by (8) who used it to explain nitrogen-limited fermentations. Our results showed that the proposed model captures the data more accurately than previously published alternatives and allowed the incorporation of macro-molecular biomass composition data (protein, mRNA and carbohydrates), which is relatively simple to obtain. The proposed mechanisms have a sound biological interpretation and were largely supported by the experimental data. In consequence, our model can be used to predict, and thus, control, *S. cerevisiae* fermentations.

2 Results

2.1 Modeling biomass formation during nitrogen-limited fermentation.

The microbial growth rate was described as the sum of two different terms. The first, μ_N, corresponds to the healthy cell division, mainly when assimilable nitrogen sources (YAN) are

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

abundant. The second, μ_C, corresponds to a secondary increase in biomass after the depletion of nitrogen sources- which we assumed corresponds, mostly to carbohydrate accumulation. The biomass growth equation reads:

$$\frac{dX}{dt} = X \cdot (\mu_N + \mu_C)$$

where X is the biomass (g/L).

To model growth associated with cell division we used a Monod type kinetics (25) where nitrogen sources were considered the limiting nutrient:

$$\mu_N = lag(t) \cdot X \cdot \mu_{maxN} \cdot \frac{N}{N + k_{sN}}$$

where μ_{max_N} is the maximum specific growth rate, N is extracellular YAN (g/L), k_{sN} is the Monod equation parameter and lag is a function of time (t) representing the lag phase. The lag phase is typically a period of adjustment in which a given microbial population adapts to a new medium before it starts growing exponentially. To model the lag we used the model proposed by (26):

$$lag = \frac{a_0}{a_0 + (1 - a_0) \cdot e^{-\mu_{maxN} \cdot t}}$$

where a_0 is an estimated parameter and μ_{maxN} is the maximum growth rate associated with μ_N.

The secondary increase in biomass concentration (μ_C) is activated by the of decrease the concentration of YAN:

$$\phi_N = (1 - \frac{N}{N + k_\phi})$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where k_{sN} is a parameter controlling the half-maximal inactivation of μ_N which is modeled with an empirical expression analogous to a proportional controller:

$$\mu_C = lag(t) \cdot \mu_{maxC} \cdot \phi_C \cdot (\theta_C - \frac{X_C}{X})$$

where θ_C is the set-point (target value) for the ratio of carbohydrates in the biomass content (X_C/X) and μ_{maxC} controls the velocity of the convergence towards that reference point.

On the contrary the formation of protein and mRNA are only affected by μ and their dynamics are described by:

$$\frac{dX_P}{dt} = \lambda_P \cdot \mu \cdot X$$
$$\frac{dX_{mRNA}}{dt} = (1 - \lambda_P - \lambda_C) \cdot \mu \cdot X$$

where λ_P corresponds the biomass content of protein.

Both M_M and M_P assume YAN consumption is proportional to μ_N:

$$\frac{dN}{dt} = -\frac{(1 - \lambda_C) \cdot \mu_N \cdot X}{Y_{X/N}}$$

where $Y_{X/N}$ is the nitrogen to mRNA and protein biomass yield.

For the sake of comparison we used two additional models: the first, regarded as M_M, corresponds to that for which the biomass follows the Monod description:

$$\frac{dX}{dt} = \mu_N$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and the second, regarded as M_C, corresponds to the model proposed by (8) and subsequently used by (9) to explain nitrogen-limited fermentations. In this model, the biomass follows a Monod equation [eq:Monod] depending on the active biomass and the dynamics of cell mass includes a death rate proportional to the extracellular ethanol concentration.

2.2 Modeling fermentation rate and production of extracellular metabolites.

Regarding fermentation rate, we proposed two models with different underlying hypotheses:

- M_P: assuming that fermentation rate is proportional to the total protein content (X_P),
- M_X: assuming that fermentation rate is proportional to the total biomass (X)

The uptake of glucose reads as follows:

$$M_P: \qquad \frac{dGlx}{dt} = -lag(t) \cdot X_P \cdot v_{Glx}$$

$$M_X: \qquad \frac{dGlx}{dt} = -lag(t) \cdot X \cdot v_{Glx}$$

where v_{Glx} is the expression governing glucose transport. A similar expression was also included for fructose (F) transport (v_F):

$$M_P: \qquad \frac{dF}{dt} = -lag(t) \cdot X_P \cdot v_F$$

$$M_X: \qquad \frac{dF}{dt} = -lag(t) \cdot X \cdot v_F$$

Following (27), we used Micheaelis-Menten type kinetics kinetics coupled to ethanol inhibition to model hexose transport:

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$v_{Glx} = v_{max,Glx} \cdot \frac{Glx}{Glx + k_{Glx}} \phi_{Eth}$$

where $v_{max,Glx}$ is the maximum rate of glucose transport, k_{Glx} is the Michaelis-Menten constant and ϕ_{Eth} ethanol inhibition. Ethanol has been reported as a non-competitive inhibitor (28) of hexose transport. Here, we modeled its effect as follows (27):

$$\phi_{Eth} = \frac{1}{1 + Eth/k_{E_i}}$$

where Eth is the extracellular ethanol concentration and k_{E_i} defines the strength of the inhibitory effect.

The excretion of extracellular metabolites was assumed to be proportional to the consumption of hexoses. In the case of ethanol, this is described by:

$$\frac{dEth}{dt} = -Y_{E/S} \cdot (\frac{dGlx}{dt} + \frac{dF}{dt})$$

where $Y_{E/S}$ is the ethanol yield produced from glucose and fructose.

2.3 Calibration of candidate models.

The Table 1 summarizes the main characteristics of the three candidate models. Additionally, a detailed description of models is given Supplementary Text S1.

The structural identifiability analysis of the models revealed that it is possible to uniquely estimate parameter values provided the dynamics of biomass, glucose uptake, nitrogen sources uptake and ethanol are measured throughout the fermentation. Remarkably, it is not necessary to

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

measure protein for the identification of the model M_P. Interested readers might find additional details in the Supplementary Text S1.

Candidate models were calibrated by data fitting using data-sets available in the literature. In particular we used data-sets taken from (13) (Exp_S) and (16) (Exp_V). The data-set Exp_S included one experiment with a nitrogen sufficient condition. The data-set Exp_V contained two experiments, with high ($Exp_V \uparrow N$) and low ($Exp_V \downarrow N$) nitrogen concentrations (50 and 300mg YAN).

We performed 10 runs of the parameter estimation problem for each case, so to guarantee convergence to the best possible solution. Since we used a metaheuristic for the optimisation of parameters, we obtained a distribution of values. We selected the best fit for ulterior analyses. Bounds for the estimated parameters values are given in supplementary text S1. Particularly, parameters related to biomass composition were chosen such that carbohydrate content does not decrease after nitrogen depletion (*i.e.* $\lambda_C > \theta_C$).

The parameters recovered with the estimation procedure are given in Supplementary Text S1. Table 2 and Figure 1 present the quality of the fits in terms of normalised root mean square error (NMRSE) for each model (M_M, M_X, M_P and M_C) and data-sets (Exp_S and Exp_V). Figures 1 A-B show that all models could explain the data within a 9% normalised root mean square error. However, the performance of different models differed substantially. The model proposed in this work, M_P, which explains total biomass dynamics using two terms (due to nitrogen and carbohydrates) produced NMRSE values that were 3.9 and 1.8 times lower than models M_M and M_C.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Since the proposed models (M_P and M_X) have more parameters than M_C and M_M, we computed the corrected Akaike Information Criterion (AIC) to rule out any possible over-fitting. We computed AIC scores excluding the residuals corresponding to protein, carbohydrates and mRNA – as these variables are not included in M_M and M_C– and considering the total number of parameters of each model (including λ_C,λ_P, K_{sC} and μ_{maxC}). Table 2 and Figures 1 C-D present the results. The model M_P presented substantial differences (< -100) with respect to M_M and M_C. Remarkably these differences are enough to guarantee that the extra parameters, used to characterise biomass composition are not inducing over-fitting and that this model is indeed the best to explain the data.

2.4 Biomass formation.

Our results suggest that Monod model is insufficient to explain biomass formation kinetics in nitrogen-limited fermentations of *S. cerevisisae*. Figure 2.A presents the NRMSE obtained for biomass in all data-sets and models. Results illustrate the proposed model M_P represents better the data in all cases. The largest differences between models were found in the Exp_S data-set. Remarkably, the normalised root mean square error obtained for the proposed model (NMSRE (M_P):1.91%) was approximately 9.7 times lower than the one obtained with the models M_M and M_C (NMSRE (M_M):18.60% and (M_C):18.56%, respectively). For the data-set Exp_V, differences were lower; still, the model (M_P) proposed in this work was always superior in the quality of fit to the biomass data.

Figure 2.B presents the fit to the time-series data for all three considered experiments. In all three cases, the proposed model could recover the increase of biomass produced after YAN depletion. The improvement was more notorious in the Exp_S data-set. Figure shows R^2 values for the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

proposed model are over 0.97 in all cases while M_M and M_C resulted in R^2 of around 0.57 for the Exp_S.

Figure 2.C shows how after nitrogen depletion, at around 20 hours (see also Figure 3.A, carbohydrates accumulate and would explain the role of μ_C in models M_P and M_X. Remarkably those models are able to recover the secondary growth observed in the biomass dynamics while M_M and M_C converge to a constant biomass value.

2.5 Fermentation rate.

All models, were able to recover the overall dynamics of hexose consumption and ethanol formation. Figure 3.A and B show the NRMSE scores for the measured variables (glucose, fructose, ethanol or glycerol) in the different experiments. Because both M_X and M_P performed well in describing biomass and share the same number of parameters, a comparison is straight-forward. Remarkably the performance of the model M_P, which uses the protein instead of the biomass to explain the fermentation rate, was always superior to that of M_x.

Figures 3.C and D show the trajectories for metabolically active cell mass for each candidate model. Models M_M and M_X do not distinguish between active and total biomass, thus curves shown in 3.C and D coincide with those in Figures 2.B). Model M_C includes a decay term that may be observed in 3.C and D. In model M_P active biomass is proportional to X_P. As shown in 3.C and D for M_X the metabolically active cell increases after YAN depletion while in M_M and M_P it stays stable; and in M_C there is an observable decay.

Figures 3.E and F show the trajectories for all candidate models and the corresponding experimental data for measured external compounds. Results show that all models perform

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reasonably well describing the dynamics of glucose, ethanol and glycerol with $R^2 > 0.8$. However, all models but M_P experiment difficulties in fitting the data Exp_V with low nitrogen. In this condition M_X (and also M_C, M_M) presented glucose consumption and ethanol productions rates faster than those observed in the experimental data; an effect not observed in M_P, indicating that protein content is a helpful biomarker while modeling fermentation rate across multiple experimental conditions.

3 Discussion

In this study, we proposed two candidate models to explain nitrogen-limited fermentations of *S. cerevisiae* in industrially controlled conditions. We proposed two modifications to the standard modelling approaches: 1) the biomass growth accounts for protein and carbohydrates, and 2) fermentation rate is proportional to the total protein content. We reconciled candidate models to previously published data.

Our results revealed that Monod model is insufficient to explain biomass formation kinetics in nitrogen-limited fermentations of *S. cerevisisae*. This is so because, during exponential growth, the biomass composition seems to remain unaltered with mRNA and protein comprising an important percentage of the biomass. However, this is not the case during secondary growth phase where μ_N fades and μ_C increases, as shown in (13) data.

Several recent studies on the modelling of alcoholic fermentation of *S. cerevisisae* relied on logistic growth (12, 20), Monod or the use of kinetic constraints coupled to a stoichiometric model (which in practice holds similar results to Monod).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Logistic growth is well suited to explain cell numbers in nitrogen-limited fermentations. However, as pointed by (25), biomass and cell numbers are not necessarily equivalent. The author argued that the average size of cells varies considerably from one phase to another in the growth cycle, thus cell concentration and bacterial density, are not equivalent. However growth rate is the same independently of being estimated in terms of one or other variable. Also in the view of (25), much confusion has been created because this important distinction has been frequently overlooked. Indeed, the author highlights that in most of the experimental problems of bacterial chemistry, metabolism, and nutrition, the significant variable is bacterial density while cell concentration is essential only in problems where division is actually concerned, or where knowledge of the elementary composition of the populations is important.

Thus, in light of our results, we argue the logistic curve model might also not adequately describe the delay during the transition from exponential growth ($\mu_N >> \mu_C$) due to carbohydrate accumulation ($\mu_C >> \mu_N$) and the different dynamics between μ_N and μ_C observed in cell mass data.

From a practical point of view, nitrogen-limited Monod models seem to explain biomass formation until the nitrogen source is depleted but struggle to explain later stages. The formulation of M_P implies that carbohydrates accumulation is solely responsible for the biomass increment after nitrogen depletion. In the case of Exp_S and $\mathrm{Exp}_V \uparrow N$, the R^2 scores for YAN, dry-weight biomass, protein and mRNA are in line with the notion that biomass increment in secondary growth phase could be explained mostly by carbohydrate accumulation.

The models proposed by (8, 9) distinguished active from inactive cells. The model M_C (adapted from (8)) was able to explain fermentation rate when data was fitted to a single experiment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

However, when the same model was applied to a data-set comprised by two different nitrogen conditions ($\mathrm{Exp}_V \downarrow N$ and $\mathrm{Exp}_V \uparrow N$) the same set of parameters had difficulties explaining hexose consumption of the low nitrogen experiment.

(20) developed a fermentation model for oenological conditions that accounted for different temperatures and initial nitrogen concentrations. These authors estimated the number of hexose transporters present in the yeast cells as a function of time, temperature and nitrogen. The model represents glucose consumption data extremely well. However, it should be noted that it comes at the cost of a heavily parameterized and nonlinear hexose transport function that also includes inhibition by substrate and ethanol. Nevertheless, the notion that hexose transporters vary with nitrogen content along the fermentation is likely to find some overlap with our simpler model for glucose transport.

(17) included the effect of ethanol inhibition on sugar transport and the existence of different hexose transporters with different maximum transport rates and substrate affinities. From a qualitative point of view, this model corresponded to a significant improvement over past models. Noticeably, the hexose transport functions considered the role of the initial nitrogen concentration and temperature. Additionally, (16) showed that the protein content of biomass is heavily dependent on the initial nitrogen concentration which supports the notion that hexose transport is likely to be better described by total protein content rather than biomass *per se*. Our model would support this hypothesis. Results obtained by M_P model are superior to those obtained with the M_X model, at least, for those cases in which initial nitrogen is restricted.

Here we proposed an alternative modelling strategy which using a minimal number of parameters, that can be estimated from easily measured data, provides a considerable

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

improvement in the description of nitrogen-limited fermentations as compared to previously published models. It should be noted that the model can be easily applied to controlled fermentations -with a single starter- provided the grape must is well characterized, i.e. YAN and carbon sources are measured beforehand and their values are followed throughout fermentation, and biomass growth and desired products are monitored over time.

4 Materials and Methods

4.1 Data

We considered data-sets from five different experiments as published in the literature. The first experiment (Exp1) data set was obtained from (13). The authors conducted a nitrogen-limited (165 mg NH_3 and 50g of glucose) batch fermentation while measuring biomass, protein content of biomass along with several extracellular metabolites (glycerol, ethanol and ammonia). We also considered four data-sets from (9). The authors performed various experiments at different conditions. In this work we denote Exp2 as that performed at 11^oC, with high-sugar and low-nitrogen; Exp3, the one performed at 15^oC, with normal sugar and low-nitrogen; Exp4, the one performed at 30^oC with normal sugar and normal nitrogen; and Exp5, the one performed at 35^oC with high-sugar and low-nitrogen.

4.2 Model building

The proposed model is composed of six ordinary differential equations (as presented in the section Results). Solutions for the system are of type:

$$y = \int_0^{t_F} \dot{y}\,(\theta, y, t)dt$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where y is the solution of the ODE system and corresponds to the dynamics of the relevant state variables - biomass, carbohydrates, protein, glucose, YAN, ethanol, glycerol- between the beginning and the end of the experiment (t_f); $\dot{y}$ represents the states time derivative.

4.3 Structural identifiability analysis

The proposed model depends on fourteen unknown parameters $\boldsymbol{\theta}$ - namely a_0, μ_{maxN}, k_{sN}, μ_{maxC}, k_{sC}, θ_C, λ_C, λ_P, $Y_{X/N}$, $v_{max,Glx}$, k_{Glx}, k_{E_i}, k_{Eth}, k_{Gly} - to be estimated from the experimental data. The proposed model introduces the description of the total protein content. However for some experimental set-ups protein was not measured. Therefore we performed a structural identifiability analysis ((29, 30)) to assess if all parameters of the model can be estimated even if protein measurements are not available.

We performed the analysis using GenSSI2 toolbox (31). The toolbox uses symbolic manipulation to implement the so-called generating series of the model and to compute identifiability *tableaus*.

4.4 Parameter estimation

The aim of parameter estimation is to compute the unknown parameters - growth related constants and kinetic parameters - that minimize the weighted least squares function which provides a measure of the distance between data and model predictions (32). In the absence of error estimates for the experimental data a fixed value for each state, the maximum experimental data, was assumed in order to normalize the least squares residuals:

$$J_{mc}(\boldsymbol{\theta}) = \sum_{k=1}^{n_{exp}} \sum_{j=1}^{n_{obs}} \sum_{i=1}^{n_{st}} \left(\frac{y_{k,j,i}(\boldsymbol{\theta}) - y_{k,j,i}^{m}}{max(y_{k,j})} \right)^2,$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

where n_{exp}, n_{obs} and n_{st} are, respectively, the number of experiments, observables, and sampling times. y_j^m represents each of the measured quantities and $y_j(\boldsymbol{\theta})$ corresponds to model predicted values.

In our particular case, we used five different experiments: the first, taken from (13), and second to fifth, taken from (9). All experiments provided time series data of the biomass (X), glucose uptake (Glx), amonia consumption (N) and ethanol production Eth.

Parameters were estimated by solving a nonlinear optimization problem to find the unknown parameter values ($\boldsymbol{\theta}$) to minimize $J_{mc}(\boldsymbol{\theta})$, subject to the system dynamics - the model- and parameter bounds (33).

To avoid the risk of premature convergence by the optimization routine while searching for the optimal parameter set, the parameter estimation procedure was repeated 10 times for each candidate model starting from different initial guesses.

4.5 Model selection

Models were compared in terms of the Akaike's information criterion (AIC). AIC compares multiple competing models (or working hypotheses) in those cases where no single model stands out as being the best. AIC is calculated using the number of data (n_d), the number fitted parameters (n_θ) and the sum of squares of the weighted residuals (WRSS). For the purpose of model comparison, we computed the AIC as follows:

$$AIC = n_d \cdot log(WRSS/n_d) + 2n_\theta + 1$$

The models were ranked by AIC, with the best approximating model being the one with the lowest AIC value. The minimum AIC value (AIC_{min}) was used to re-scale the Akaike's

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

information criterion. The re-scaled value $\Delta AIC = |AIC_i - AIC_{min}|$ was used to assess the relative merit of each model. Models such as $\Delta \leq 2$ have substantial support, models for which $4 \leq \Delta \leq 7$ have considerably less support and models with $\Delta > 10$ have no support (34).

4.6 Numerical methods and tools

The parameter estimation and model selection were implemented in the AMIGO2 toolbox (35). The system of ODEs was compiled to speed up calculations and solved using the CVODES solver (36), a variable-step, variable-order Adams-Bashforth-Moulton method. The parameter estimation problem was solved using a hybrid meta-heuristic, the enhanced scatter search method (eSS, (37)).

5 Supplemental Material

All scripts and data required to reproduce results and figures can be accessed in https://sites.google.com/site/amigo2toolbox/examples.

6 Acknowledgments

We thank the three anonymous reviewers for their comments on the original version of the manuscript. This project has received funding from MCIU/AEI/FEDER, UE (grant reference RTI2018-093744-B-C33) and GAIN Xunta de Galicia (Contrato Programa and grant reference IN607B 2020-03).

1. Parapouli M, Vasileiadis A, Afendra AS, Hatziloukas E. 2020. *Saccharomyces cerevisiae* and its industrial applications. AIMS Microbiol 6:1–31.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2. Bisson LF. 1999. Stuck and sluggish fermentations. Am J Enol Viticult 50:107–119.

3. Alexandre H, Charpentier C. 1998. Biochemical aspects of stuck and sluggish fermentation in grape must. J Ind Microbiol Biotech 20:20–27.

4. Tesnière C, Brice C, Blondin B. 2015. Responses of *saccharomyces cerevisiae* to nitrogen starvation in wine alcoholic fermentation. Appl Microbiol Biotechnol 99:7025–7034.

5. Bell S-J, Henschke PA. 2005. Implications of nitrogen nutrition for grapes, fermentation and wine. Aust J Grape Wine Res 11:242–295.

6. Marín MR. 1999. Alcoholic fermentation modelling: Current state and perspectives. Am J Enol Viticult 50:166–178.

7. Miller KV, Block DE. 2020. A review of wine fermentation process modeling. J Food Eng 273:109783.

8. Cramer AC, Vlassides S, Block DE. 2002. Kinetic model for nitrogen-limited wine fermentations. Biotechnol Bioeng 77:49–60.

9. Coleman MC, Fish R, Block DE. 2007. Temperature-dependent kinetic model for nitrogen-limited wine fermentations. Appl Environ Microbiol 73:5875–5884.

10. Kelkar S, Dolan K. 2012. Modeling the effects of initial nitrogen content and temperature on fermentation kinetics of hard cider. J Food Eng 109:588–596.

11. David R, Dochain D, Mouret J-R, Wouwer AV, Sablayrolles J-M. 2010. Dynamical modeling of alcoholic fermentation and its link with nitrogen consumption. IFAC Proceedings Volumes 43:496–501.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12. Henriques D, Alonso-del-Real J, Querol A, Balsa-Canto E. 2018. *Saccharomyces cerevisiae* and *s. Kudriavzevii* synthetic wine fermentation performance dissected by predictive modeling. Front Microbiol 9:88.

13. Schulze U, Liden G, Villadsen J. 1996. Dynamics of ammonia uptake in nitrogen limited anaerobic cultures of saccharomyces cerevisiae. J Biotechnol 46:33–42.

14. Jørgensen H, Olsson L, Rønnow B, Palmqvist E. 2002. Fed-batch cultivation of baker's yeast followed by nitrogen or carbon starvation: Effects on fermentative capacity and content of trehalose and glycogen. Appl Microbiol Biotechnol 59:310–317.

15. Thomsson E, Gustafsson L, Larsson C. 2005. Starvation response of *saccharomyces cerevisiae* grown in anaerobic nitrogen-or carbon-limited chemostat cultures. Appl Environ Microbiol 71:3007–3013.

16. Varela C, Pizarro F, Agosin E. 2004. Biomass content governs fermentation rate in nitrogen-deficient wine musts. Appl Environ Microbiol 70:3392–3400.

17. Pizarro F, Varela C, Martabit C, Bruno C, Pérez-Correa JR, Agosin E. 2007. Coupling kinetic expressions and metabolic networks for predicting wine fermentations. Biotechnol Bioeng 98:986–998.

18. Vargas FA, Pizarro F, Pérez-Correa JR, Agosin E. 2011. Expanding a dynamic flux balance model of yeast fermentation to genome-scale. BMC Syst Biol 5:75.

19. Sainz J, Pizarro F, Pérez-Correa JR, Agosin E. 2003. Modeling of yeast metabolism and process dynamics in batch fermentation. Biotechnol Bioeng 81:818–828.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

20. Malherbe S, Fromion V, Hilgert N, Sablayrolles J-M. 2004. Modeling the effects of assimilable nitrogen and temperature on fermentation kinetics in enological conditions. Biotechnol Bioeng 86:261–272.

21. Charnomordic B, David R, Dochain D, Hilgert N, Mouret J-R, Sablayrolles J-M, Vande Wouwer A. 2010. Two modelling approaches of winemaking: First principle and metab. eng. Math Comput Model Dyn Syst 16:535–553.

22. David R, Dochain D, Mouret J-R, Wouwer AV, Sablayrolles J-M. 2014. Nitrogen-backboned modeling of wine-making in standard and nitrogen-added fermentations. Bioproc Biosyst Eng 37:5–16.

23. Mouret J-R, Farines V, Sablayrolles J-M, Trelea I-C. 2015. Prediction of the production kinetics of the main fermentative aromas in winemaking fermentations. Biochem Eng J 103:211–218.

24. Perez M, Luyten K, Michel R, Riou C, Blondin B. 2005. Analysis of *saccharomyces cerevisiae* hexose carrier expression during wine fermentation: Both low-and high-affinity hxt transporters are expressed. FEMS Yeast Res 5:351–361.

25. Monod J. 1949. The growth of bacterial cultures. Annu Rev Microbiol 3:371–394.

26. Baranyi J, Roberts TA. 1994. A dynamic approach to predicting bacterial growth in food. Int J Food Microbiol 23:277–294.

27. Hjersted JL, Henson MA, Mahadevan R. 2007. Genome-scale analysis of *saccharomyces cerevisiae* metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97:1190–1204.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

28. Leão C, Van Uden N. 1984. Effects of ethanol and other alkanols on passive proton influx in the yeast *saccharomyces cerevisiae*. Biochim Biophys Acta Biomembr 774:43–48.

29. Chis O-T, Banga JR, Balsa-Canto E. 2011. Structural identifiability of systems biology models: A critical comparison of methods. PLoS ONE 6.

30. Chis O-T, Villaverde AF, Banga JR, Balsa-Canto E. 2016. On the relationship between sloppiness and identifiability. Mathematical Biosciences 282:147–161.

31. Ligon TS, Fröhlich F, Chiş OT, Banga JR, Balsa-Canto E, Hasenauer J. 2018. GenSSI 2.0: Multi-experiment structural identifiability analysis of SBML models. Bioinformatics 34:1421–1423.

32. Walter E, Pronzato L. 1996. On the identifiability and distinguishability of nonlinear parametric models. Mathematics and computers in simulation 42:125–134.

33. Vilas C, Arias-Méndez A, García MR, Alonso AA, Balsa-Canto E. 2018. Toward predictive food process models: A protocol for parameter estimation. Crit Rev Food Sci Nutr 58:436–449.

34. Burnham KP, Anderson R. 2004. Multimodel inference: Understanding AIC and BIC in model selection. Sociol Meth & Res 33:261–304.

35. Balsa-Canto E, Henriques D, Gábor A, Banga JR. 2016. AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology. Bioinformatics 32:3357–3359.

36. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, Woodward CS. 2005. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Trans Math Softw(TOMS) 31:363–396.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

37. Egea JA, Balsa-Canto E, García M-SG, Banga JR. 2009. Dynamic optimization of nonlinear processes with an enhanced scatter search method. Ind Eng Chem Res 48:4388–4401.

Table1. Main characteristics of candidate models. *The complete mathematical formulation, including all parameters and their units are reported in the Supplementary Text S1.*

Model identifier	Model characteristics
M_P	State variables: Biomass, YAN, Glucose, Ethanol,Glycerol, Protein,mRNA
	Biomass is split into two terms, due to nitrogen and carbohydrates
	Lag-phase
	Fermentation rate is proportional to the total protein content
M_X	State variables: Biomass, YAN, Glucose, Ethanol,Glycerol, Protein,mRNA
	Biomass is split into two terms, due to nitrogen and carbohydrates
	Lag-phase
	Fermentation rate is proportional to the biomass
M_M	State variables: Biomass, YAN, Glucose, Ethanol, Glycerol
	Biomass follows Monod, being nitrogen the limiting substrate
	Lag-phase
	Fermentation rate is proportional to the biomass
M_C	State variables: Active Biomass, YAN, Glucose, Ethanol, Glycerol
	Lag-phase (not included in the original formulation)
	Fermentation rate is proportional to the active biomass

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table2. Summary of candidate models characteristics and quality of fit scores. *Observed variables regard the variables included in each model. N_{Data} regards the amount of data used to compute the quality of fit scores. N_{Pars} regards the number of parameters to be fitted. N_{RMSE} regards the normalised mean root squared error, AIC regards the corrected Akaike criterion and ΔAIC the re-scaled Akaike. The table shows the model M_P results in the minimum N_{RMSE} and AIC values. Attending to the re-scaled Akaike criterion, M_X is the closest model to the best. However, attending to* 34*, models with* $|\Delta AIC| > 10$*, have no support.*

ExpS; Schulze et al. 1996

	MM	MX	MP	MC
NData	122	122	122	122
NPars	9	14	14	10
NRMSE	0.08	0.02	0.02	0.08
AIC	-574.64	-886.88	-899.61	-572.71
ΔAIC	**-324.97**	**-12.73**	**0**	**-326.9**

ExpV; Varela et al. 2004

	MM	MX	MP	MC
NData	99	99	99	99
NPars	10	16	16	11
NRMSE	0.05	0.04	0.02	0.05
AIC	-564.74	-597.38	-672.47	-564.52
ΔAIC	**-107.73**	**-75.09**	**0**	**-107.95**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

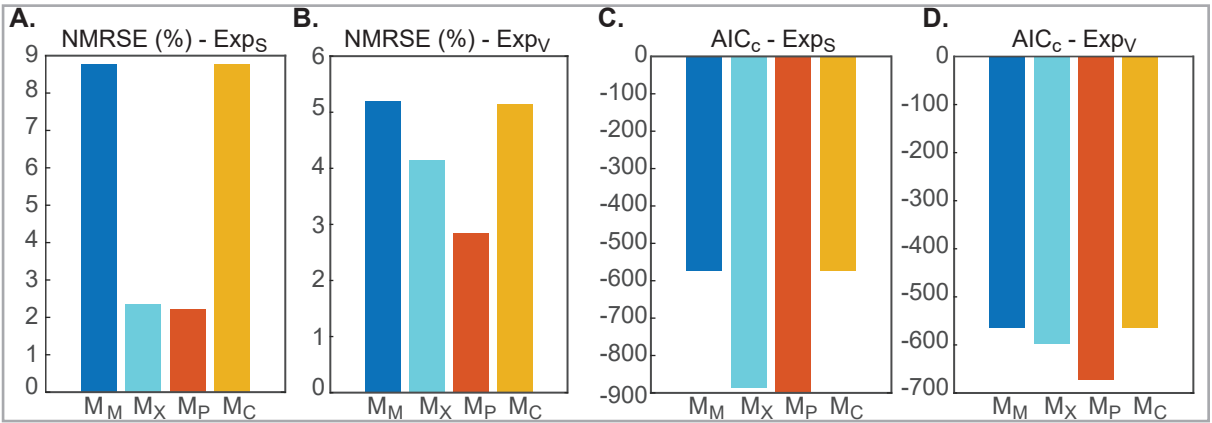
Figure 1. Comparison of candidate models. *A-B Show the normalised root mean square error (NRMSE in %) o C-D present the corrected Akaike Information Criterion (AIC_c). Values correspond to all models and the two experimental data sets Exp_S and Exp_V. For the case of AIC_c only YAN, glucose, fructose, ethanol and glycerol residuals were compared, since those are the variables present in all models. Results show how the proposed model M_P is superior in terms of both scores for both experimental data sets, differences are particularly important in the fit to Exp_V in which high and low nitrogen fermentations are fitted simultaneously.*

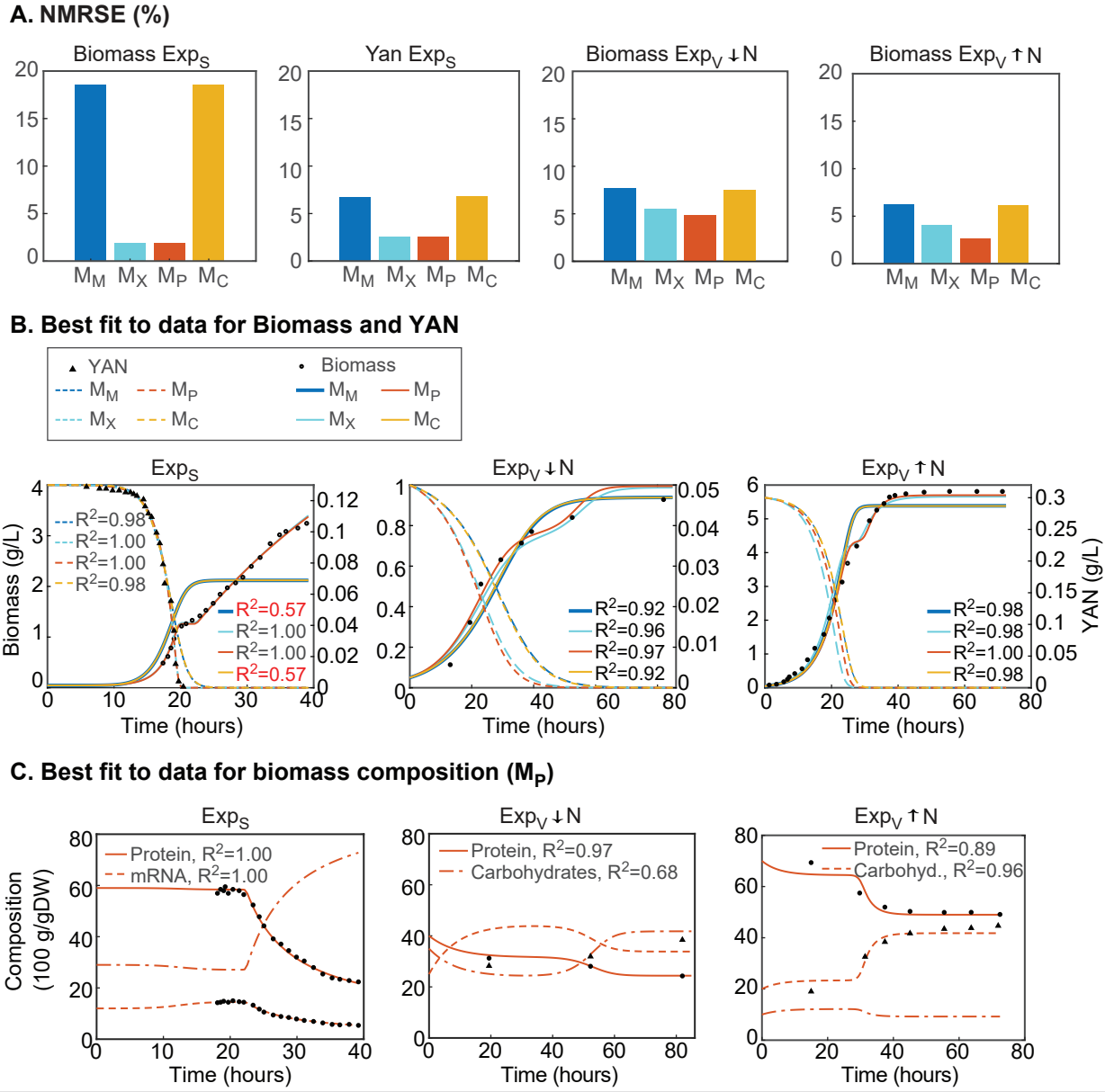
Figure 2. Comparison of candidate models in the fitting of biomass data. *Figure A presents the normalised root mean square error of dry-weight biomass and YAN in all candidate models and experiments. Figure B presents the time-course trajectories and data for YAN and dry-weight biomass in all experiments. Figure C presents the time-course trajectories as obtained by model M_P of the percentage of protein ($100 \cdot X_P/X$), mRNA ($100 \cdot X_{mRNA}/X$) and carbohydrates ($100 \cdot X_C/X$) present in the biomass. Data-set Exp_V is comprised by experiments $Exp_V \uparrow N$ and $Exp_V \downarrow N$ which were fitted in combination and share the same parameters. In contrast Exp_S corresponds to a single experiment fitted in isolation. Model M_P, proposed here, results in better fits for dry-weight biomass in terms of NRMSE and AIC_c. The model can recover the increase in biomass observed in several experiments after YAN depletion.*

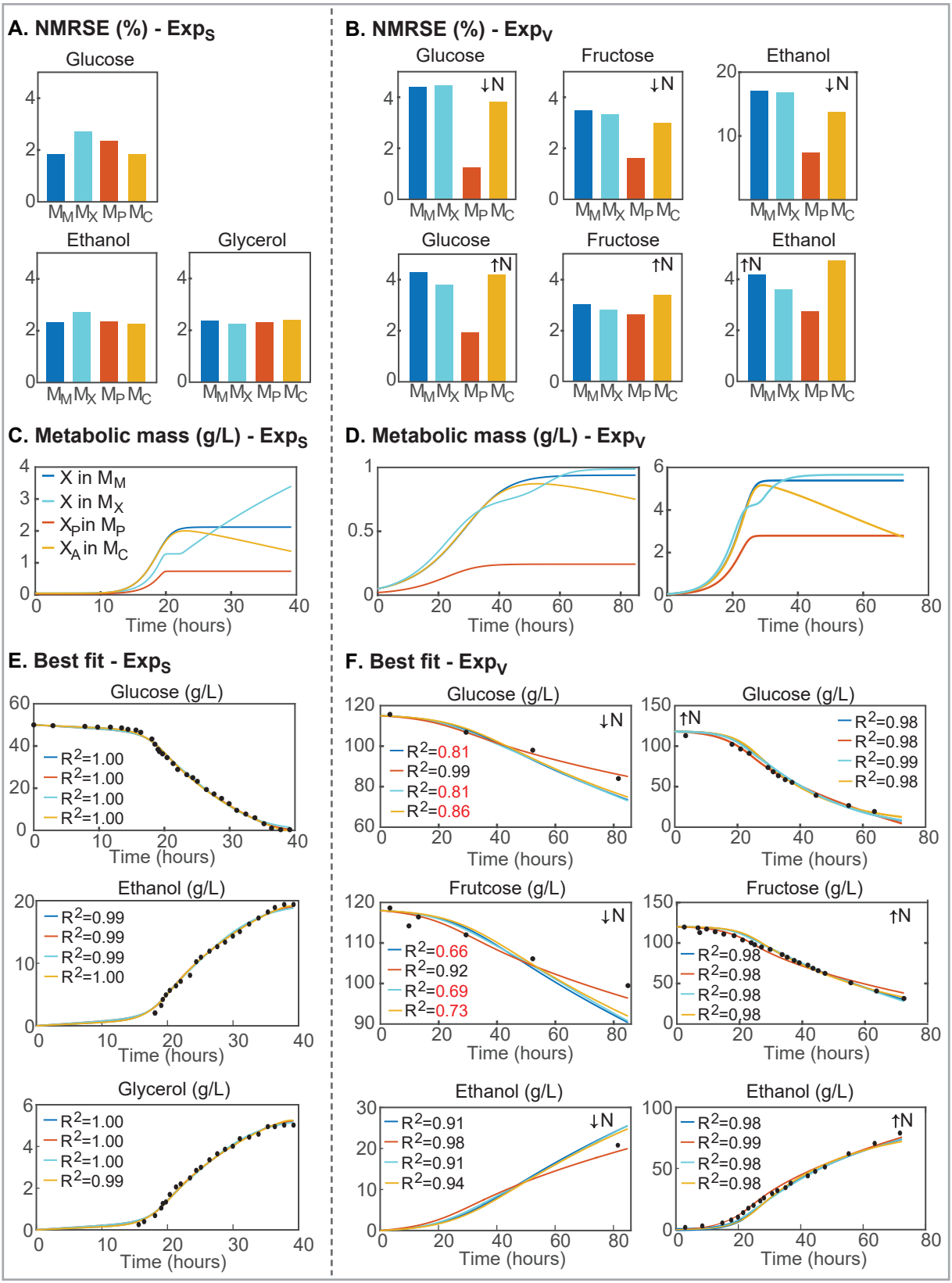
Figure 3. Comparison of candidate models in the fitting of fermentation data. *Figures A and B present the NRMSE for the measured external compounds for the different experiments and models; Figures C and D present the time-course trajectories for the metabolically active cell*

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.09.447824; this version posted June 10, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mass (X, X_A or X_P); Figures E and F present the time-course trajectories of the measured variables in all experiments. R^2 values are reported for each observable; those values below 0.9 *are marked in red. Note that all models are quite successful in fitting the experimental data for Exp_S while higher differences are observed for experiments Exp_V. Only, M_P resulted in R^2 values above* 0.9 *for all states and conditions.*


A.
NMRSE (%) - Exp$_S$
9
8
7
6
5
4
3
2
1
0
M_M M_X M_P M_C
B.
NMRSE (%) - Exp$_V$
6
5
4
3
2
1
0
M_M M_X M_P M_C
C.
AIC$_c$ - Exp$_S$
0
-100
-200
-300
-400
-500
-600
-700
-800
-900
M_M M_X M_P M_C
D.
AIC$_c$ - Exp$_V$
0
-100
-200
-300
-400
-500
-600
-700
M_M M_X M_P M_C


A. NMRSE (%)
Biomass Exp$_S$
Yan Exp$_S$
Biomass Exp$_V$ ↓N
Biomass Exp$_V$ ↑N
M$_M$ M$_X$ M$_P$ M$_C$
B. Best fit to data for Biomass and YAN
YAN
Biomass
Exp$_S$
Exp$_V$ ↓N
Exp$_V$ ↑N
Biomass (g/L)
YAN (g/L)
Time (hours)
R^2=0.98
R^2=1.00
R^2=1.00
R^2=0.98
R^2=0.57
R^2=1.00
R^2=1.00
R^2=0.57
R^2=0.92
R^2=0.96
R^2=0.97
R^2=0.92
R^2=0.98
R^2=0.98
R^2=1.00
R^2=0.98
C. Best fit to data for biomass composition (M$_P$)
Composition (100 g/gDW)
Protein, R^2=1.00
mRNA, R^2=1.00
Protein, R^2=0.97
Carbohydrates, R^2=0.68
Protein, R^2=0.89
Carbohyd., R^2=0.96


A. NMRSE (%) - Exp$_S$
Glucose
Ethanol
Glycerol
B. NMRSE (%) - Exp$_V$
Glucose
Fructose
Ethanol
↓N
↑N
M_M M_X M_P M_C
C. Metabolic mass (g/L) - Exp$_S$
X in M_M
X in M_X
X_P in M_P
X_A in M_C
Time (hours)
D. Metabolic mass (g/L) - Exp$_V$
Time (hours)
E. Best fit - Exp$_S$
Glucose (g/L)
R^2=1.00
R^2=1.00
R^2=1.00
R^2=1.00
Ethanol (g/L)
R^2=0.99
R^2=0.99
R^2=0.99
R^2=1.00
Glycerol (g/L)
R^2=1.00
R^2=1.00
R^2=1.00
R^2=0.99
Time (hours)
F. Best fit - Exp$_V$
Glucose (g/L)
↓N
R^2=0.81
R^2=0.99
R^2=0.81
R^2=0.86
Glucose (g/L)
↑N
R^2=0.98
R^2=0.98
R^2=0.99
R^2=0.98
Frutcose (g/L)
↓N
R^2=0.66
R^2=0.92
R^2=0.69
R^2=0.73
Fructose (g/L)
↑N
R^2=0.98
R^2=0.98
R^2=0.98
R^2=0.98
Ethanol (g/L)
↓N
R^2=0.91
R^2=0.98
R^2=0.91
R^2=0.94
Ethanol (g/L)
↑N
R^2=0.98
R^2=0.99
R^2=0.98
R^2=0.98
Time (hours)